A Steady Performance Stopping Criterion for Pareto-based Evolutionary Algorithms

Olga Roudenko, Marc Schoenauer

A Steady Performance Stopping Criterion for Pareto-based Evolutionary Algorithms

Olga Rudenko[1], Marc Schoenauer[2]

[1] Centre de Mathématiques Appliquées, Ecole Polytechnique,
91128 Palaiseau Cedex, France
roudenko@cmapx.polytechnique.fr

[2] INRIA, team TAO
LRI, bat. 490, Université Paris-Sud
91405 Orsay Cedex, France
Marc.Schoenauer@inria.fr

Abstract. The most commonly used stopping criterion in Evolutionary Multi-objective Algorithms is an a priori fixed number of generations (or evaluations). But it is rather difficult to speak about achieving a particular compromise between the quality of the final solutions and the computation time when stopping an algorithm this way. Unfortunately, whereas single-objective Evolutionary Algorithms can stop when the fitness does not improve during a given number of generations, such "steady-fitness" stopping criterion does not easily extend to the multi-objective framework.
This paper introduces a stability measure based on the density of the non-dominated solutions and proposes to use it to stop the optimization process when no significant improvement is likely to take place on further iterations. This approach is validated by the empirical results obtained applying NSGA-II to the well-known bi-objective ZDT-benchmarks. In particular, the problem ZDT4 best illustrates the ability of the proposed criterion to avoid useless continuation of a wedged optimization process when a local Pareto-optimal set is reached.

Introduction

When talking about "improving the performance of a numerical method", one generally means "obtaining better results at the smallest possible computation cost for the largest possible class of problems". In that sense, the choice of the stopping criterion does not look like being able to improve the performance of an optimization algorithm: it can neither increase its convergence speed, nor improve its accuracy. However, when applying the algorithm to a real world problem, the stopping criterion becomes an important factor of its *practical efficiency*. In every particular case, the efficiency is evaluated with respect to the specification of the experts of the application domain. In particular, most specifications contain constraints and preferences defining desirable compromises between the quality of the solution and the computation time: an appropriately chosen stopping condition may be very important in order to achieve such compromise.

Evolutionary Algorithms (EAs) are becoming more and more popular to solve difficult real-world optimization problems that resist traditional methods. Moreover, specific EAs, called EMAs (Evolutionary Multi-objective Algorithms), have been developed for multi-objective problems [2], whose output is a sampling of an approximation of the Pareto front of the problem at hand. However, no satisfactory stopping criterion does exist for EMAs.

For single-objective EAs, there are three basic stopping criteria:

1. stop after a fixed number of iterations
2. stop when a pre-defined optimization function value is attained
3. stop when a fixed number of iterations are performed without improvement

In practice, two or three of them are often combined, usually criterion 1 with one of the others (or both of them), to ensure a finite computation time.

In EMA framework, however, only criterion 1 above is used. Indeed, EMA Darwinian stages (selection and replacement, see Section 1.1) are based on the relation of Pareto dominance, that provides the main criterion for the comparison of the individuals (see Section 1). This means that the "fitness" of each individual depends on the whole population: it is hence impossible to use something like criterion 2 above based on fitness.

In the present paper, we propose an adaptation of criterion 3 above for EMAs. We claim that the ability to stop an algorithm when no more significant improvement can reasonably be expected is very important for its successful application to any real world problem in order to avoid wasting precious computation time.

Let us note that this study is entirely based on the NSGA-II method presented in the section 1.3.

The paper is organized as follows. The first section introduces the evolutionary multi-objective optimizers, in particular, the algorithms based on the notion of the Pareto dominance. Section 2 discusses some specific features of NSGA-II, one of the best performing EMA to-date: the proposed "steady performance" stopping criterion is at the moment limited to NSGA-II, and is based on the stabilization of the density measure used by NSGA-II to preserve diversity. Section 3 presents some experimental results illustrating the applicability of the proposed stopping condition. It is followed by the conclusions and indicates some directions for further research.

1 Evolutionary Multi-objective Optimization

Multi-objective optimization aims at simultaneously optimizing several contradictory objectives. For such kind of problems, there does not exist a single optimal solution, and compromises have to be made.

An element of the search space x is said to *Pareto-dominate* another element y if x is not worse than y with respect to all objectives, and is strictly better than y with respect to at least one objective. The set of all elements of the search space that are not Pareto-dominated by any other its element is called the *Pareto set* of the multi-objective problem at hand: it represents the best possible compromises with respect to the contradictory objectives.

Solving a multi-objective problem amounts to choose one solution among those non-dominated solutions, and some decision arguments have to be given. Unlike classical optimization methods, that generally find one of the Pareto optimal solutions by making the initial optimization problem single-objective, EMA are to-date the only algorithms that actually directly search for the whole Pareto front, allowing decision makers to choose one of the Pareto solutions with more complete information.

1.1 Evolutionary Algorithms

This subsection is especially targeted toward the readers who are not familiar with Evolutionary Optimization, it contains the basic notions that are necessary for understanding the remainder of the paper. For more details, the interested reader is invited to read the recent book by Eiben and Smith [6].

Crudely mimicking the Darwinian evolution of natural populations, based on *natural selection* and *blind variation operators*, EAs evolve a "population of individuals", i.e. a set of N elements of the search space.

Starting from an initial randomly-initialized population, the basic iteration loop of an EA (be it single- of multi-objective) is (see figure 1): SELECT some parents for reproduction; apply VARIATION OPERATORS (e. g. crossover, mutation) to those selected parents to generate offspring ; REPLACE some parents with some offspring.

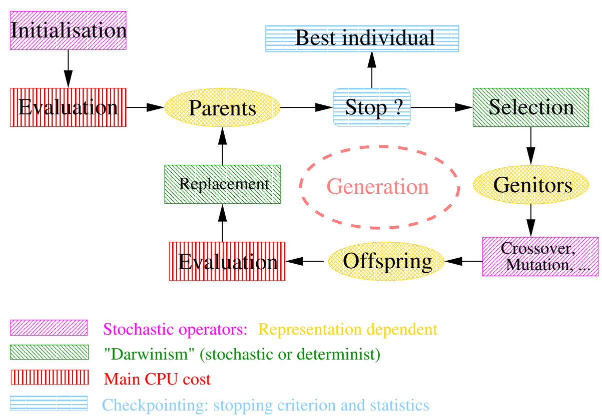
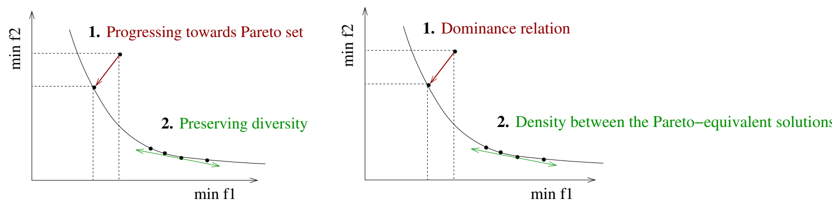


Fig. 1. Standard EA loop

The (only) difference between single-objective and multi-objective EAs lies in the Darwinian phases, selection (choice of the individuals that will reproduce) and replacement (choice of the individuals that will survive to the next generation). While in single-objective EAs, Darwinian selection is straightforwardly based on the value of the objective, EMAs consider Pareto dominance as the most important criterion to compare individuals. Let us note however that the Pareto dominance relation establishes only a partial order among the individuals.

1.2 Pareto-based Evolutionary Algorithms

In order to find a good approximation of the Pareto set (uniform and well spread sampling of the non-dominated solutions sufficiently close to the Pareto set of the problem at hand), EMAs have to enforce some progress toward the Pareto front while at the same time preserving diversity between the non-dominated solutions (figure 2(a)).



(a) *EMA goals* **(b)** *Criteria to compare the individuals*

Fig. 2. EMA goals and corresponding criteria for the comparison of the individuals

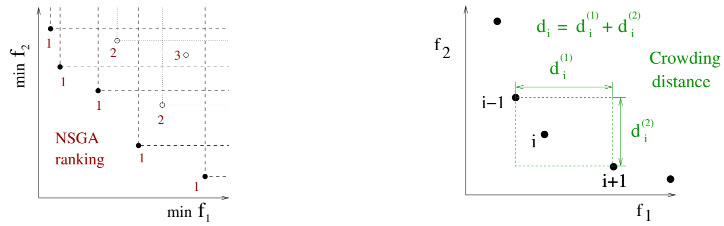
Numerous evolutionary methods have recently been designed for the particular task of searching for the Pareto set (the interested reader will find a good summary in [2]). The best performing among them (NSGA-II [3], SPEA2 [9], PESA [1]) are directly based on the notion of Pareto dominance (figure 2(b)). Among diversity preserving techniques, some were transposed to EMAs from single-objective EAs (such as sharing, for instance), while others, like the crowding distance that will be described in next subsection, have been developed within EMA framework.

Elitism has been recognized as another important feature of EMAs [8]. The notion of elitism, in the EMA framework, is directly related to the notion of the Pareto dominance: the non-dominated individuals can be preserved either by maintaining an archive (SPEA2 and PESA) or by using a deterministic replacement procedure (NSGA-II).

1.3 NSGA-II

The NSGA-II algorithm has been proposed by Deb et al. in 2001 [3]. The progress toward the Pareto set is here due to the *Pareto ranking* that divides the population into non-dominated subsets, as illustrated by figure 3(a): first, all non-dominated individuals of the population are labeled as being of rank 1 ; then they are temporarily removed from the population and the process is repeated: the non-dominated individuals of the remainder of the population are given rank 2, and so on, until the whole population is ranked.

The diversity preserving technique is based on the *crowding distance* - one of the possible estimations of the density of the solutions belonging to the same non-dominated subset. The crowding distance of each individual i is computed as follows: the non-dominated subset to which the individual i belongs is ordered following each of the objectives; for each objective m, the distance $d_i^{(m)} = f_m(i+1) - f_m(i-1)$ between both neighbors of the individual i according to objective m is computed (Fig. 3(b)); the sum of these distances over all objectives gives the crowding distance



(a) *Ensuring progress toward the Pareto set* **(b)** *Preserving diversity technique*

Fig. 3. NSGA-II criteria to compare the individuals

of the individual i (the average of all $d_i^{(m)}$ over $m = 1, ..., M$ can also be used).

According to the general formulation of EMA criteria for the comparison of individuals (see figure 2(b)), the following comparison operator is used at the Darwinian stages of NSGA-II:

1. Rank(x) < Rank(y) => x is better then y
2. Rank(x) = Rank(y)
   crowd_dist(x) > crowd_dist(y) => x is better then y

NSGA-II selection is based on tournaments i.e. to choose an individual for reproduction, T individuals (generally, $T = 2$) are randomly picked from the population and compared to each other using the comparison operator defined above. The winner then becomes a genitor (Fig. 1). NSGA-II replacement is deterministic: it consists in merging all parents and offspring (see Figure 1), and choosing the N best individuals in that global population using the same comparison operator. The algorithm NSGA-II is said to be elitist because the best non-dominated individuals are preserved from one generation to another (though this issue will be discussed in greater detail in section 5).

2 Stopping an "inefficient" evolution

Maximizing the efficiency ratio *quality/calculation_time* has generally the highest priority when solving a real world numerical problem. Clearly, its value decreases if numerous iterations are performed with a very small (or without any) improvement of the quality of the solution. In terms of Evolutionary Optimization, if such a situation persists for too long, the evolution is stuck and should better be interrupted. Note that stagnating in an evolutionary optimization process (in the single-objective as well as in the multi-objective case) does not necessarily correspond to being very close to the optimum. It is however important to enable the detection of such "hopeless" situations in order to avoid wasting computation time performing useless iterations. The detection of stagnation, in the absence of any knowledge about how far the search is from the optimum, can be considered as a signal to stop, eventually trying to get better solutions by some another way (e.g. restarting another evolution, or applying the local search).

In this section, we propose a criterion to detect the moment when NSGA-II evolution can be considered stuck. This criterion can easily be reformulated for other Pareto-based EMAs, but, at the moment, its applicability was only tested with NSGA-II. This study is based on the observations made for bi-objective benchmark problems ZDT1-ZDT4 [8] for which it is easy to observe the population dynamics in the objective space. At first, the evolution was considered stuck when it became difficult to visually distinguish different generations in the objective space.

2.1 Dynamics of NSGA-II

In this section, we shall take a close look at the behavior of the population with respect to Pareto dominance during a run of NSGA-II.

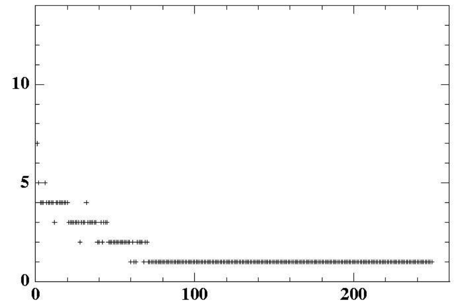
The disappearance of all dominated individuals When observing the dynamics of the NSGA-II populations, it has been remarked that as the population gets closer to the Pareto surface, the progress of the population in the direction improving all the criteria slows down and the number of different Pareto ranks quickly decreases to one (see Figure 4(a)), i.e. all the individuals become Pareto-equivalent.

This disappearance of the dominated individuals comes from the replacement mechanism of NSGA-II, based on the comparison operator that gives priority to the individuals of the smaller rank. But it can also be noticed that, as long as the population has enough space to progress, it does not "enter" into such state of Pareto-equivalent population. According to Deb [2], the deceleration of the optimization process during the late evolution stage is actually due to the absence of dominated individuals in the population i.e. to the lack of diversity in the Pareto-dominance direction. On our opinion, however, both these phenomena (the "non-dominated population" as well as the deceleration of the progress) are explained by the proximity of a Pareto-optimal set (be it global or local) that makes finding the new dominating solutions less and less likely.

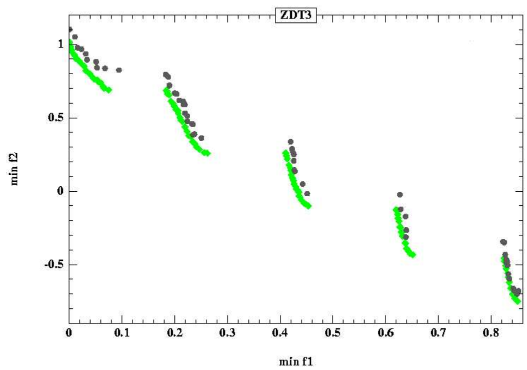
Nevertheless, as illustrated for instance by figure 4(b), even after the whole population has become Pareto-equivalent, some meaningful progress toward the Pareto set can still be made. It is clear that the moment when the whole population is of rank 1 is not the right time to stop the evolution.

The deterioration As already mentioned in section 1.3, the NSGA-II is an elitist EMA. Taking this fact in account, one could think about a stopping criterion based on the number of newly created non-dominated individuals that dominate at least one previously non-dominated individual : as this number is expected to decrease when the search nears the Pareto optima, it could be seen as a possible measure of improvement. However, a closer look at the dynamics of this quantity makes clear that no particular tendency can be noticed : it continuously and endlessly varies between zero and approximately 20% of the population size.

However, this phenomenon is not as strange as it could seem at first sight, and is far from witnessing some infinite improvement. This comes from the fact that EMA



(a) *Generations* vs *Number of the Pareto ranks*



(b) Snapshots: 70th (black points) and 100th (gray points) generations

Fig. 4. *The dominated individuals disappear (forever) at generation 70; nevertheless, visually noticeable improvement is still made later on (b).*

elitism is based on the Pareto dominance relation which provides only relative performance measure. Indeed, in spite of this elitism, the replacement mechanism of NSGA-II does not prevent the deterioration [7]: because of the the diversity-maintaining technique, some non-dominated individuals can be removed from the population and, at some later stage, an individual can appear which would have been dominated by this previously eliminated individual but which is non-dominated in the current population. Figure 5 illustrates such a situation. Moreover, as the populations near some Pareto optimal set (local or global), the deterioration becomes more and more frequent.

The limit of sampling uniformity In NSGA-II method, the uniformity of the sampling of non-dominated solutions is measured by the crowding distances (section 1.3). Observing the dynamics of the crowding distances of all rank-1 individuals, we notice a clear tendency toward stabilization : see the behavior of the maximal, minimal and average crowding distance values with respect to generations on figure 6(a) – the same run than that of figure 4. It has been repeatedly noticed that the average distance typically stabilizes when the whole population becomes Pareto-equivalent (e.g.

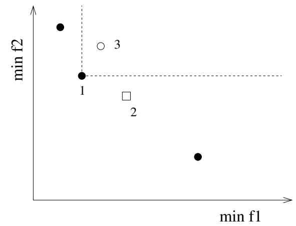
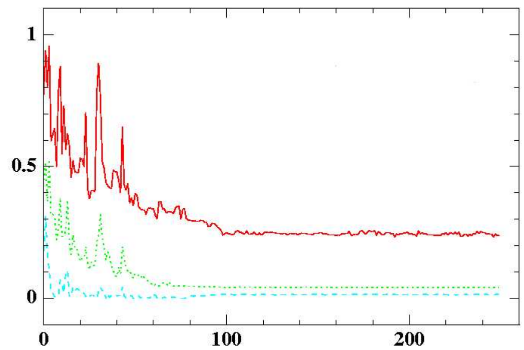
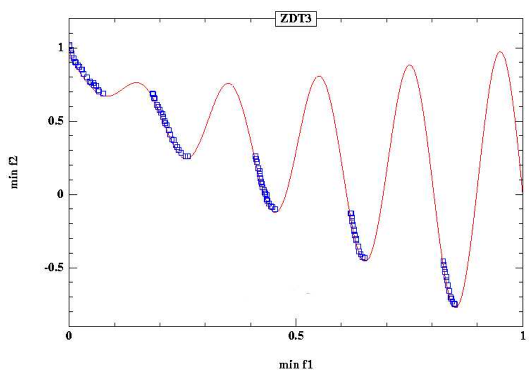


Fig. 5. *Deterioration in NSGA-II: on the generation N, the point 1 is replaced by the point 2 ; on the generation N+1 the point 3 may be kept as it will be non-dominated*



(a) Max, Average and Min distances dynamics



(b) 100th generation

Fig. 6. *On the 100th generation the population is sufficiently close to the Pareto front (non-dominated part of the curve) to allow ignoring further improvement*

at generation 70 in Figure. 4 and in Figure 6(a)). Moreover, in all our observations, the maximal crowding distance stabilizes last, after the minimal and average (Fig. 6(a)). Furthermore, simultaneous observing the dynamics of the maximal crowding distance, and the corresponding behavior of the population in the objective space, shows that the stabilization of the maximal distance approximately takes place when the last "gaps" in the non-dominated solutions distribution disappear. The progress in the "dominance direction" is at that moment already practically indistinguishable.

To sum up, no significant improvement has been observed after the stabilization of the maximal crowding distance, neither in the sense of the progress in the "Pareto direction", nor in the sense of the sampling uniformity. This observation supports the idea of stopping NSGA-II when the maximal crowding distance stabilizes.

2.2 A stability measure

The implementation of the stopping condition based on the stabilization of the maximal crowding distance requires, first of all, to choose an appropriate stability measure.

Let us denote d_l maximal crowding distance computed at generation l. One possible way to measure its stability over L generations is to simply calculate the difference

$$\delta_L = max_{l=1}^{L} d_l - min_{l=1}^{L} d_l.$$

In that case, the stopping condition would be written as follows:

$$\delta_L < \delta_{lim}. \tag{1}$$

But in fact, some oscillations of d_l are still possible after its stabilization, hence making the appropriate choice of δ_{lim} difficult : if δ_{lim} is small, condition (1) is difficult to satisfy ; if δ_{lim} is sufficiently large to take in account possible oscillations, then there is a risk that the algorithm stops much too early.

Hence, in order to smooth out the effect of such oscillations, we propose to use the standard deviations rather than δ_L :

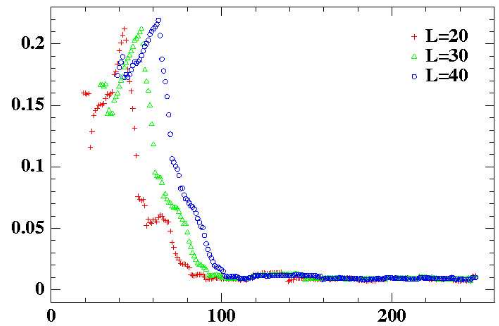
$$\sigma_L \stackrel{\text{def.}}{\equiv} \sqrt{\frac{1}{L}\sum_{l=1}^{L}(d_l - \overline{d}_L)^2} < \delta_{lim}, \tag{2}$$

where $\overline{d}_L$ is the average of d_l over L generations. Note that the value of σ_L does not depend on the actual values of the objective functions, as all crowding distances are normalized.

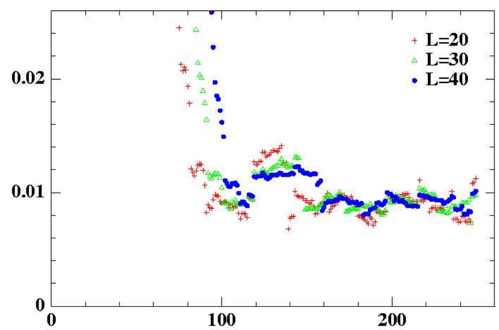
2.3 Hints for choosing the parameters

Two parameters need to be defined before the stopping condition (2) can be used: L, the length of the time window, and δ_{lim}. We will give here some general suggestions about their choice. These suggestions are based on the observations of the dynamics of σ_L.

Clearly, σ_L stabilize as a consequence of the stabilization of d_l. In fact, the quantities σ_L (for different values of L) stabilize at the same level (see figure 7) that we denote $\sigma_{L,stab}$. The value of δ_{lim} must be chosen slightly bigger then $\sigma_{L,stab}$. Indeed, if $\delta_{lim} \approx \sigma_{L,stab}$, there is a high risk to miss the right moment to stop the algorithm. On the other hand, again, a too large value of δ_{lim} may lead to some premature stopping.



(a) General view



(b) Zoom

Fig. 7. σ_L *dynamics*

By the way, it has been noticed that $\sigma_{L,stab}$ depends on the population size N : $\sigma_{L,stab}$ increases as N decreases. Thus, the value of δ_{lim} has to be adjusted when changing N.

Regarding the choice of L, it is important to take in account the fact that, the smaller L, the larger the oscillation of σ_L, both before (see Fig. 7(a)) and after (see Fig. 7(b)) stabilization. Hence, a too small value for L could once again lead to premature stopping.

3 Experimental results

As already mentioned, experiments have been continuously performed during the tuning of the proposed stopping criterion. This section summarizes the results of applying our stopping criterion on the four benchmarks ZDT functions (see [2]).

In all experiments described in this section, δ_{lim} and L have been chosen so that they ensure stopping the algorithm rather a bit too late then a bit too early for all the four benchmarks ZDT1-ZDT4. The following common parameters have been used for all experiments: population size set to 100, $L = 40, \quad \delta_{lim} = 0.02$. This means that, for each NSGA-II run, the algorithm was stopped when the following inequality became true:

$$\sigma_{40} \leq 0.02. \tag{3}$$

3.1 Stopping when converged

Figure 8 shows the non-dominated solutions obtained after 21 runs of NSGA-II stopped using criterion (3). The benchmarks ZDT1-ZDT3 simulate different difficulties related to the shape of the Pareto front: ZDT1 has a convex front (Fig. 3(a)), ZDT2 has a concave front (Fig. 3(b)), and ZDT3 has a discontinue front (Fig. 3(c)). For all three problems, the algorithm finds, at every run, a very good approximation of the exact Pareto front.

Let us notice that allowing 40 iterations to detect stability ($L = 40$) is quite a lot considering the proportion of the overall average generation number (statistics given in figure 8). But choosing a smaller value of L with the same δ_{lim} would create an additional risk of premature stopping for some NSGA-II runs, in particular for ZDT2 and ZDT3. Moreover, if δ_{lim} is simultaneously reduced, it becomes too close to the value of $\sigma_{L,stab}$, and we might miss the appropriate moment to stop. When as many as 21 runs can be performed, a reasonable reduction of L is not expected to significantly influence the overall quality of the solutions. But in practice, the optimal computation cost (very often defined by the number of evaluations) corresponds to the smallest number of runs, even if each of them needs a little more iterations to find better solutions with higher probability : this is ensured by choosing L sufficiently large.

3.2 Interruption when wedged

Problem ZDT4 was designed to exhibit numerous local Pareto fronts. The presence of such local fronts very often prevents the convergence of NSGA-II (at least in our implementation) toward the exact Pareto front of the problem ZDT4 (Figure 9).

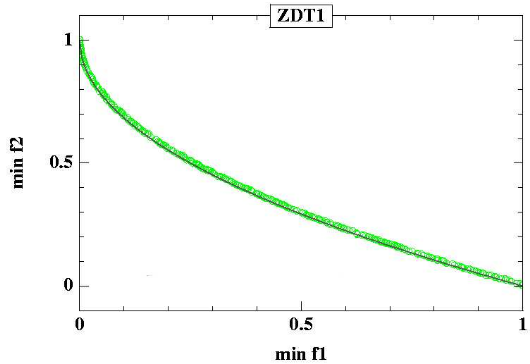
Observing the populations dynamics, we notice that, in all of our experiments, NSGA-II indeed never overcomes the "obstacle" after it is wedged at some local Pareto front. In such situation, using the stopping condition 2 corresponds preventing useless iterations. This can be the appropriate moment to apply, for example, some local search that would conclude the optimization process.

The detection of the wedged evolution, by checking condition (3), is illustrated in the figures 10 and 11.

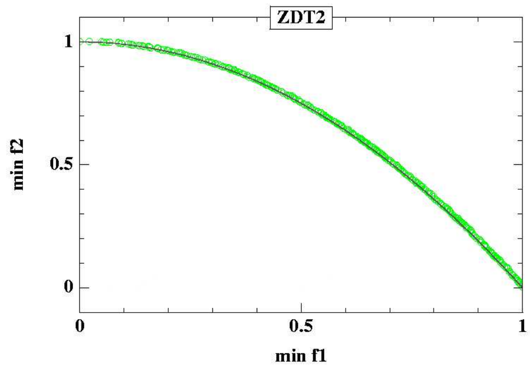
4 Conclusions and future work

The stopping criterion for EMAs proposed in this paper is an analogous of the "steady fitness" stopping condition used in the single-objective EAs. It is based on the observations of the NSGA-II population dynamics, and has been tested on four commonly used bi-objective test problems from EMO literature.

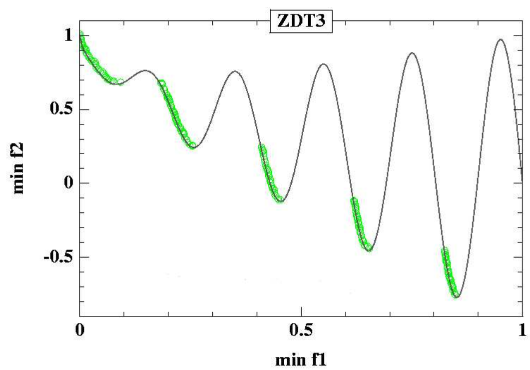
Some suggestions concerning the choice of the parameters have been given. In particular, it has been established that the parameter values should depend on the



(a) *94 - 98 - 103 generations over 21 runs*



(b) *106 - 116 - 132 generations over 21 runs*



(c) *97 - 110 - 133 generations over 21 runs*

Fig. 8. Non-dominated solutions obtained by NSGA-II over 21 runs, for problems ZDT1, ZDT2 and ZDT3 (gray points) compared to the exact Pareto fronts (black curves) ; basic statistics (minimum, average and maximum over all runs) of the number of generations performed until condition (3) was met are given under the corresponding figures.

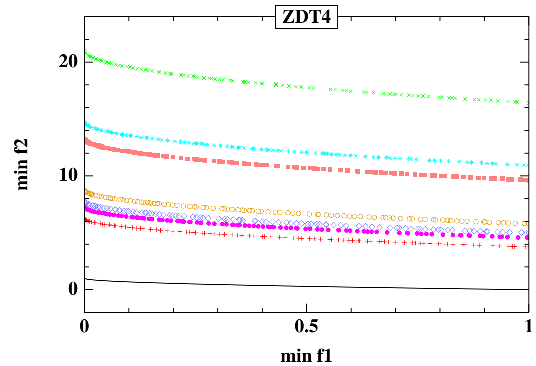
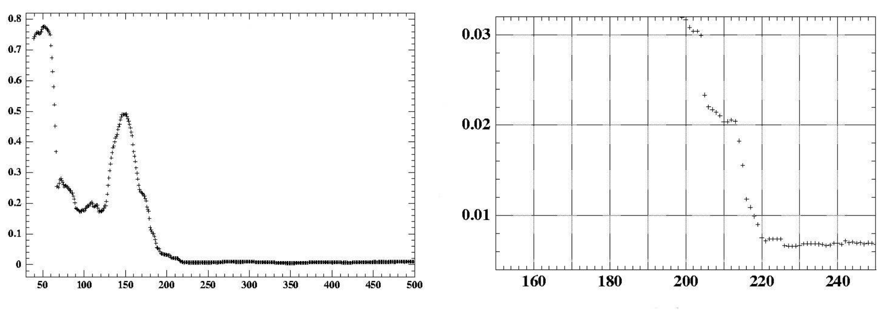


Fig. 9. *Non-dominated solutions after 7 different NSGA-II runs compared to the exact ZDT4 Pareto front (black curve): the convergence to the local Pareto fronts takes place*



(a) σ_{40} dynamics

(b) σ_{40} dynamics (zoom)

Fig. 10. *According to the stopping criterion (3) the present NSGA-II run would be stopped at generation 214.*

population size and, in some sense, on the number of the consecutive runs that can be performed for every particular application.

Note that some observations that have been made for NSGA-II may not stand for other EMAs. On the other hand, it has been noticed [5] that the diversity preserving technique of NSGA-II does not scale up when the number of the objectives exceeds two. Hence, the next step of research will be to adapt the proposed stopping condition to SPEA2, and to test its applicability for problems with 3-4 objectives. Let us notice, however, that in such cases, the "visual" analysis of the populations dynamics will not be possible, and we plan to apply the so-called *running metrics* [4] in order to evaluate the applicability of our EMA "steady performance" stopping condition and to learn about the appropriated choice of parameters.

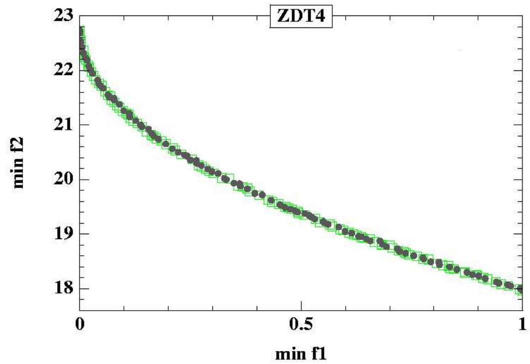


Fig. 11. *Populations at generation 214 (black points) and 500 (hardly distinguishable gray squares) illustrate the absence of any improvement after the population is "trapped" into a local Pareto front.*

References

1. D. W. Corne, J. D. Knowles, and M.L. Oates. The pareto envelope-based selection algorithm for multiobjective optimization. In M. Schoenauer and al., editors, *Proceedings of the* 6^{th} *Conference on Parallel Problem Solving from Nature*, pages 839–848. Springer Verlag, 2000.
2. K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. New York: John Wiley, 2001.
3. K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In M. Schoenauer and al., editors, *Proceedings of the* 6^{th} *Conference on Parallel Problem Solving from Nature*, pages 849–858. Springer Verlag, 2000.
4. K. Deb and S. Jain. Running performance metrics for evolutionary multi-objective optimization. Technical Report 2002004, Indian Institute of Technology, Kanpur, May 2002.
5. K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable test problems for evolutionary multi-objective optimization. Technical Report 2001001, Indian Institute of Technology, Kanpur, 2001.
6. A. E. Eiben and J. E. Smith. *Introduction to Evolutionary Comuting*. Springer Verlag, 2003.
7. M. Laumanns, L. Thiele, K. Deb, and E. Zitzler. Combining convergence and diversity in evolutionary multiobjective optimization. *Evolutionary Computation*, 10(3):263–282, 2002.
8. E. Zitzler, K. Deb, and L. Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation Journal*, 8(2):125–148, 2000.
9. E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the strength pareto evolutionary algorithm. Technical Report 103, Computer Engineering and Networks Laboratory, ETH, Zurich, Switzerland, 2001.